

Datenbereinigung

statistik-online.ch

Nach dem Export von Daten aus Umfrageprogrammen wie LimeSurvey, Qualtrics oder Google Forms müssen die Daten häufig zuerst bereinigt werden. Ziel der Datenbereinigung ist es, fehlerhafte, unvollständige oder unbrauchbare Daten zu erkennen und die Daten für statistische Analysen vorzubereiten.

1 Typische Schritte der Datenbereinigung

1. Unnötige Variablen entfernen

Viele Umfrageprogramme exportieren zusätzliche technische Variablen, die für die Analyse nicht benötigt werden.

Beispiele:

- Startdatum
- Enddatum
- IP-Adresse
- Antwortdauer
- Vorschauseiten
- Interne IDs

Diese Variablen können meist gelöscht werden, um den Datensatz übersichtlicher zu machen.

2. Fehlende Werte prüfen

Überprüfen Sie, wie viele fehlende Werte (Missing Values) vorhanden sind.

Wichtige Fragen:

- Gibt es Personen mit sehr vielen fehlenden Antworten?
- Gibt es Variablen mit ungewöhnlich vielen fehlenden Werten?
- Sollten bestimmte Personen ausgeschlossen werden?

Beispiel:

- Personen mit ausschliesslich fehlenden Werten löschen
- Personen mit weniger als 50% beantworteten Fragen ausschliessen

3. Wertebereiche kontrollieren

Prüfen Sie, ob unmögliche oder fehlerhafte Werte vorhanden sind.

Beispiele:

- Alter = 250 Jahre
- Zustimmungsskala von 1–5 enthält plötzlich den Wert 9
- Negative Werte bei Körpergewicht

Häufig entstehen solche Fehler durch falsche Eingaben oder fehlerhafte Kodierungen.

4. Datentypen überprüfen

Kontrollieren Sie, ob Variablen im richtigen Format gespeichert wurden.

Beispiele:

- Zahlen wurden als Text gespeichert
- Dezimalzahlen wurden mit Komma statt Punkt eingegeben
- Offene Antworten enthalten Zahlen und Text gleichzeitig

Beispiel:

- Statt „25“ wurde „fünfundzwanzig“ eingegeben
- Statt „180“ wurde „ca. 180 cm“ eingegeben

Solche Werte müssen überprüft und gegebenenfalls korrigiert oder als fehlend codiert werden.

5. Antwortmuster prüfen

Manche Personen beantworten Umfragen nicht sorgfältig.

Mögliche Hinweise:

- Immer dieselbe Antwort gewählt
- Extrem kurze Bearbeitungszeit
- Widersprüchliche Angaben

Beispiele:

- Alter = 12 Jahre, aber „Masterabschluss“
- Alle Antworten einer Skala = „5“

6. Kodierungen kontrollieren

Antwortkategorien müssen häufig numerisch codiert werden.

Beispiel:

Antwort	Kodierung
stimme überhaupt nicht zu	1
stimme eher nicht zu	2
neutral	3
stimme eher zu	4
stimme voll zu	5

Achten Sie darauf, dass die Kodierung korrekt und konsistent ist.

7. Items umpolen

Bei Fragebögen enthalten manche Items eine umgekehrte Bedeutung.

Beispiel: Skala zu Stress mit Codierung 1 - 5

- Item: „Ich bin oft gestresst“
- Item: „Ich fühle mich meistens entspannt“

Vor der Berechnung von Skalenwerten müsste das zweite Item umgepolt werden, damit in beiden Items ein höherer Wert für höheren Stress steht.

Original	Umpolung
1	5
2	4
3	3
4	2
5	1

8. Reliabilität prüfen (Cronbachs Alpha)

Bei Skalen mit mehreren Items sollte überprüft werden, ob die Items zuverlässig zusammenpassen.

Cronbachs Alpha (Interne Konsistenz):

- $\alpha \geq .90$ = ausgezeichnet
- $\alpha \geq .80$ = gut
- $\alpha \geq .70$ = akzeptabel
- $\alpha \geq .60$ = kritisch prüfen

Ein niedriges Alpha kann darauf hinweisen, dass:

- Items nicht zusammenpassen
- Items falsch kodiert wurden
- ein Item vergessen wurde umzupolen

9. Skalenwerte berechnen

Nach der Datenbereinigung können Skalenwerte erstellt werden.

Typische Möglichkeiten:

- Mittelwert der Items einer Skala
- Summe der Items einer Skala

Beispiel:

- Stress-Skala aus 5 Fragen
- Berechnung eines Mittelwerts pro Person

10. Dokumentation der Änderungen

Alle Änderungen am Datensatz sollten dokumentiert werden, damit die Datenaufbereitung nachvollziehbar bleibt.

Beispiele:

- Welche Personen ausgeschlossen wurden
- Welche Variablen gelöscht wurden
- Welche Items umgepolt wurden

2 Zusammenfassung

Schritt	Ziel
Unnötige Variablen entfernen	Datensatz vereinfachen
Fehlende Werte prüfen	Unvollständige Daten erkennen
Wertebereiche kontrollieren	Fehlerhafte Werte finden
Datentypen überprüfen	Falsche Formate erkennen
Antwortmuster prüfen	Unzuverlässige Antworten erkennen
Kodierungen kontrollieren	Einheitliche Werte sicherstellen
Items umpolen	Richtige Richtung der Skala herstellen
Cronbachs Alpha prüfen	Reliabilität der Skala prüfen
Skalenwerte berechnen	Analyse vorbereiten
Änderungen Dokumentieren	Nachvollziehbare Datenaufbereitung